ZKG INTERNATIONAL

ISSN: 2366-1313

IDENTIFICATION OF FAKE FACEBOOK PROFILE USING ARTIFICIAL NEURAL NETWORK

[1]M. JHANSI RANI, [2]N SRI BHUVAN, [3]A. MANEESHA, [4]C. AKSHAY

[1]Assistant Professor, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad, jhansirani512@gmail.com

[2]BTech student, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad, nsribhuvan@gmail.com

[3]BTech student, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad, ankurimaneesha@gmail.com

[4]BTech student, Dept.of IT, TKR College of Engineering & Technology, Meerpet, Hyderabad,

chandamakshy08@gmail.com

Abstract: *Fake profiles play an important role in advanced persisted threats and are also involved in other malicious activities. We know present Social Networks plays an important role for internet users to carry out their daily activities. At the same time different kinds of scammers are also equally attracted towards these social media. These guys are creating fake profiles to spread their content and carryout for scams. In this project, we use deep learning, namely an artificial neural network to determine what are the chances that Facebook account details are authentic or not. In this project, we have taken the Facebook profile Dataset from GitHub to identify the fake profiles. We also outline the classes and libraries involved. Furthermore, we discuss the sigmoid function and how the weights are determined and used. Finally, we consider the parameters of the social network page which are utmost important in the provided solution. The other dangers of personal data being obtained for fraudulent purposes is the presence of bots and fake profiles. Bots are programs that can gather information about the user without the user even knowing. This process is known as web scraping. What is worse, is that this action is legal. Bots can be hidden or come in the form of a fake friend request on a social network site to gain access to private information.*

Keywords: *Identification of fake facebook profiles, artificial neural networks, social networks.*

I. INTRODUCTION

In 2017, the Facebook population reached two. Forty-six billion users, they have become the most popular desire on social networks. Social networks earn income from the information provided by users. The average consumer now does not know that his rights have been lost by simply using the social network service. Social media agencies have a lot to gain from consumer pricing. Every time a user shares a new area, new photos, likes, dislikes and tags other users on the posted content, Facebook generates revenue through ads and news feeds. More specifically, the average American consumer generates about $26.76 per region. This number dries up quickly as hundreds of thousands of users participate. In today's digital age, increasing reliance on information technology has left the average citizen vulnerable to crime along with statistical leaks and possible identity theft. These attacks can happen without warning and often without notifying the victims of the factual abuse. At the moment, there is little incentive for social networks to improve the security of their statistics. These breaches often target social networks such as Facebook and Twitter. They can also target banks and other monetary institutions. There seems to be a news issue about hacking on social media every single day. Facebook recently suffered a statistic breach that affected about 50 million customers. Facebook offers a really specific set of provisions that define what they do with user records. Coverage does nothing or does nothing to prevent continued exploitation of your security and privacy. Fake profiles seem to slip through Facebook's built-in security features.

Various risks of getting personal data for fraudulent jobs are the presence of bots and fake profiles. Bots are programs that can collect information about a user without the user understanding it. This technique is known as web scraping. The worst thing is that this movement is a prison. Bots can be hidden or come in the form of a fake friend request on a social media page to gain access to non-public information.

The answer given in this challenge aims to learn about the dangers of a bot in the form of a fake profile on your social networks. This solution can come in the form of a set of rules. The language we choose for the application is Python. The set of rules may be able to determine whether the current friend request a person receives online is a real person, a bot, or a fake friend request looking for information. Our set of rules can work with the help of social media groups because we will need an educational data set from them to teach our version and then determine if the profiles are fake or not. The ruleset should function as a traditional layer in a consumer's web browser as a browser plug-in.

MOTIVATION

Online Social Networks are a great venue for scammers to impersonate the identities of users via creating fake profiles. Fake profiles are a popular tool for the intruders which can be used to carry out malicious activities such as impersonation attacks and harming person’s reputation and privacy in Online Social Networks. Hence, recognizing the identities of fake profiles is one of the critical security problems in Online Social Networks.

PROBLEM IDENTIFICATION

Malicious users create fake profiles to phish login information from unsuspecting users. A fake profile will send friend requests to many users with public profiles. These counterfeit profiles bait unsuspecting users with pictures of people that are considered attractive. Once the user accepts the request, the owner of the phony profile will spam friend requests to anyone this user is a friend. The fake profile's contents typically have links that lead to an external website where the damage happens. An unaware curious user clicking the bad link will damage their computer. The cost can be as simple as catching a virus to as bad as installing a rootkit turning the computer into a zombie. While Facebook has a rigorous screening to keep these fake accounts out, it only takes one fake profile to damage the computers of many.

Long range interpersonal communication has end up a notable diversion inside the web as of now, drawing in countless clients, burning through billions of minutes on such administrations. Online Social organization (OSN) administrations assortment from social cooperations based stages like Facebook or MySpace, to understanding spread driven stages suggestive of twitter or Google Buzz, to social association trademark brought to introduce frameworks like Flicker. The contrary hand, improving security concerns and safeguarding the OSN privateness actually connote a most significant bottleneck and saw mission. While utilizing social organization's (Sn's), exceptional people share stand-out amounts of their private arrangement. Having our singular expertise totally or to some degree uncovered to the overall population, makes us amazing focuses for exceptional kinds of attacks, the most exceedingly terrible of which could be ID burglary.

Data fraud happens when any singular uses character's skill for a private achieve or reason. During the prior years, online recognizable proof burglary has been an essential issue thinking of it as impacted huge number of individuals around the world. Casualties of ID robbery might experience remarkable sorts of punishments; for delineation, they would lose time/cash, get dispatched to reformatory, get their public picture destroyed, or have their associations with partners and friends and family harmed. As of now, by far most of SN's does no longer checks common users‟ obligations and has entirely vulnerable privateness and security approaches. Truth be told, most SN's applications default their settings to negligible privateness; and subsequently, SN's turned into a best stage for misrepresentation and misuse. Person to person communication contributions have worked with data fraud and Impersonation assaults for genuine comparable to innocent assailants.

To compound the situation, clients are expected to outfit right comprehension to set up a record in Social Networking sites. Simple observing of what clients share on-line would prompt devastating misfortunes, not to mention, assuming such bills had been hacked. Profile data in web-based organizations will likewise be static or dynamic. The subtleties which can be provided with the guide of the individual on the hour of profile creation is known as static information, the spot as the important part that are described with the guide of the framework inside the organization is called dynamic information. Static information incorporates segment components of an individual and his/her advantages and dynamic information remembers individual runtime propensities and region for the organization. By far most of momentum research relies upon static and dynamic information. Anyway, this isn't applicable to heaps of the informal organizations, where handiest some of static profiles are seen and dynamic profiles as a rule are not clear to the individual organization. In excess of a couple of methods have been proposed by exceptional specialist to understand the phony characters and malignant substance material in web-based informal communities. Each interaction had its own merits and negative marks. The issues including long range interpersonal communication like security, web-based tormenting, abuse, and savaging and numerous others.

II. LITERATUE SURVEY

In this project, we use deep learning, namely an artificial neural network to determine what are the chances that Facebook friend request is authentic or not. In this project, we have taken the Facebook profile Dataset from GitHub to identify the fake profiles. We also outline the classes and libraries involved. Furthermore, we discuss the sigmoid function and how the weights are determined and used. Finally, we consider the parameters of the social network page which are utmost important in the provided solution.

• Sybil rank was designed in late 2012, to efficiently identify fake profiles through a ranking graph-based system. The algorithm uses a seed selection method combined with early terminated random walks to propagate trust. Its computational cost is measured in O (nlogn). Profiles are ranked according to the number of interactions, tags, wall posts, and friends over time. Profiles that have a high rank are considered to be real with fake profiles having a low rank in the system. Unfortunately, this technique was found to be mostly unreliable because it failed to take into account the possibility that real profiles can be ranked low and fake profiles can be ranked high.

• Sarod and Mishra proposed a different approach which is a sequence of steps to detect fake profiles . the Facebook graph API tool to gain access to numerous profiles and wrotea script to extract the viewed information. Later on, this extracted information forms the attributes the classifier will use in their algorithm. First, the data is in JSON format, which is further parsed to a structured format (CSV) that is easier readable by machine learning techniques. These commas separated values will later make the classifier more efficient. The authors tried unsupervised and also supervised machine learning techniques. In this case, supervised machine learning techniques had a higher accuracy rate of almost 98%. For supervised machine learning, they split up the dataset into training and testing sets. They used80% of the samples to train the classifier and the rest to test it. After the algorithm runs, there is feedback provided to the profile, requiring it to submit identification to prove it is not a fakeprofile.

• Profiles are processed on mass to extract features. Resilient Back Propagation algorithm in neural networks algorithm combined with support vector machines is used in the classification of fake profiles.

- Sybil Frame uses multi-stage level classification. Approaches include content-based and structure based. Content-based approach explores the dataset and extracts information used to calculate prior information about nodes and edges. Structure-based approach correlates nodes using Markov random field and loopy belief propagation which employs previous information. The content-based approach is used in the first stage of Sybil Frame and Structure-based approach is used in the second stage of Sybil Frame technique.

- Clickstreams are analysed, and Friend recommendations are examined in stage I. Vote Trust uses a voting-based system that pulls user activities to find fake profiles using trust-based vote assignment and global votes total. It is considered as the first line of defines due to limitations which include real accounts that were already compromised being sold.

III. PROPOSED SYSTEM

In this Project we are using Artificial Neural Networks to identify whether given social network account details are from genuine or fake users. ANN algorithm will be trained with all previous users fake and genuine account dataset and then whenever we gave new test data then that ANN train model will be applied on new test data to identify whether given new account details are from genuine or fake users.

Online social networks such as Facebook or Twitter contains users details and some malicious users will hack social network database to steal or breach users information, To protect users data we are using ANN Algorithm.

To train ANN algorithm we are using below details from social networks

Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP, Status

All fake user’s main intention is to send friend request to normal users to hack their machine or to stealtheir data and never they will have many numbers of posts or have many following friends and their account age also will have a smaller number of years. By this features Facebook willmark whether user profile is fake or genuine. This Facebook profile data we downloaded from Facebook website and using this data to train ANN model. Below are some values from profile dataset.

Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP, Status

10, 1, 22, 0, 1073, 237, 0, 0, 0

10, 0, 33, 0, 127, 152, 0, 0, 0

10, 1, 46, 0, 1601, 405, 0, 0, 0

10, 0, 25, 0, 704, 380, 0, 0, 0

7, 1, 34, 1, 64, 721, 1, 1, 1

7, 1, 30, 1, 69, 587, 1, 1, 1

7, 1, 36, 1, 61, 782, 1, 1, 1

7, 1, 52, 1, 96, 827, 1, 1, 1

In above dataset all bold names are the dataset column names and all integer values are the dataset values. As ANN will not take string value so we convert gender values to 0 or 1, if male value is 1 and if female value is 0. In above dataset last column give us information of fake or genuine account if last column contains value 0 then account is genuine otherwise fake. All fake account will have less number of posts as their main intention is to send friend requests not posts, so by analysing this features Facebook mark that record with value 1 which means it's a fake account. We are using above dataset to train ANN model and this dataset saved inside code 'dataset' folder. After building train model we input test data with account details and ANN will give resultas fake or genuine. Below are some values from test data.

Account_Age, Gender, User_Age, Link_Desc, Status_Count, Friend_Count, Location, Location_IP

10, 1, 44, 0, 280, 1273, 0, 0

10, 0, 54, 0, 5237, 241, 0, 0

7, 0, 42, 1, 57, 631, 1, 1

7, 1, 56, 1, 66, 623, 1, 1

In above test data STATUS column and its value is not there and ANN will predict status and give us result whether above test data is fake or genuine. In output we can see result of above test data as genuine or fake.

Framework For Identification of Fake Profiles

The sequence of processes that need to be followed for continues detection of fake profiles with active leaning from the feedback of the result given by the classification algorithm. This framework can easily be implemented by the social networking companies.

1. The detection process starts with the selection of the profile that needs to be tested.

2. After selection of the profile, the suitable attributes (i.e. features) are

selected on which the classification algorithm is implemented.

3. The attributes extracted is passed to the trained classifier. The classifier gets trained regularly as new training data is feed into the classifier.

4. The classifier determines the whether the profile is fake or genuine.

5. The classifier may not be 100% accurate in classifying the profile so; the feedback of the result is given back to the classifier.

6. This process repeats and as the time proceeds, the no. oftraining data increases and the classifier becomes more and more accurate in predicting the fake profiles.

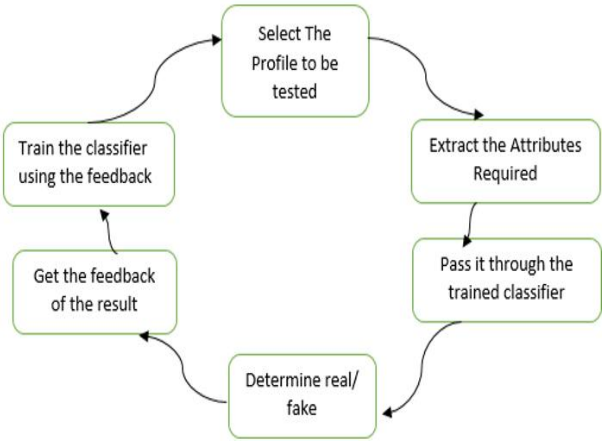

Select The Profile to be tested
Extract the Attributes Required
Pass it through the trained classifier
Determine real/ fake
Get the feedback of the result
Train the classifier using the feedback

Fig.1 Framework for identification of fake profiles

Proposed Algorithms

To demonstrate how to build an ANN neural network-based image classifier, we shall build a 6-layer neural network that will identify and separate one image from other. This network that we shall build is a very small network that we can run on a CPU as well. Traditional neural networks that are very good at doing image classification have many more parameters and take a lot of time if trained on normal CPU. However, our objective is to show how to build a real-world convolutional neural network using TENSORFLOW.

Neural Networks are essentially mathematical models to solve an optimization problem. They are made of neurons, the basic computation unit of neural networks. A neuron takes an input (say x), do some computation on it (say: multiply it with a variable w and adds another variable b) to produce a value (say; z= wx + b). This value is passed to a non-linear function called activation function (f) to produce the final output (activation) of a neuron. There are many kinds of activation functions. One of the popular activation functions is Sigmoid.

The neuron which uses sigmoid function as an activation function will be called sigmoid neuron. Depending on the activation functions, neurons are named and there are many kinds of them like

RELU, Tan H. If you stack neurons in a single line, it’s called a layer; which is the next building block of neural networks.

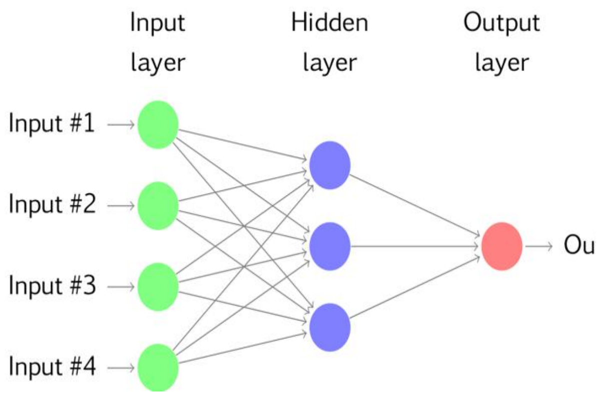


Fig.2 Structure of Artificial Neural Networks

HOW DOES ARTIFICIAL NEURAL NETWORK WORKS?

To predict class label multiple layers operate on each other to get best match layer and this process continues till no more improvement left.

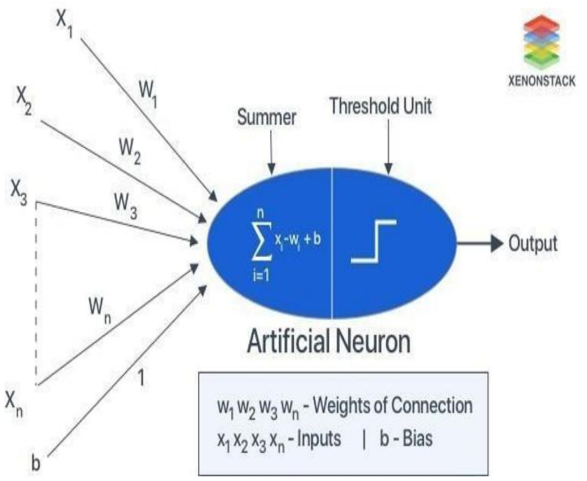

X_1
X_2
X_3
X_n
b
W_1
W_2
W_3
W_n
1
Summer
Threshold Unit
XENONSTACK
$\sum_{i=1}^{n} x_i w_i + b$
Output
Artificial Neuron
$W_1 W_2 W_3 W_n$ - Weights of Connection
$X_1 X_2 X_3 X_n$ - Inputs | b - Bias

Fig.3 Working of ANN

Artificial Neural Networks can be viewed as weighted directed graphs in which artificial neurons are nodes, and directed edges with weights are connections between neuron outputs and neuron inputs.

• The Artificial Neural Network receives information from the external world in pattern and image in vector form. These inputs are designated by the notation x(n) for n number of inputs.

• Each input is multiplied by its corresponding weights. Weights are the information used by the neural network to solve a problem. Typically, weight represents the strength of the interconnection between neurons inside the Neural Network.

• The weighted inputs are all summed up inside the computing unit (artificial neuron). In case the weighted sum is zero, bias is added to make the output not- zero or to scale up the system response. Bias has the weight and input always equal to ‘1'.

• The sum corresponds to any numerical value ranging from 0 to infinity. To limit the response to arrive at the desired value, the threshold value is set up. For this, the sum is forward through an activation function.

- The activation function is set to the transfer function to get the desired output. There are linear as well as the nonlinear activation function.

What are the commonly used activation functions?

Some of the commonly used activation function is - binary, sigmoidal (linear) and tan hyperbolic sigmoidal functions(nonlinear).

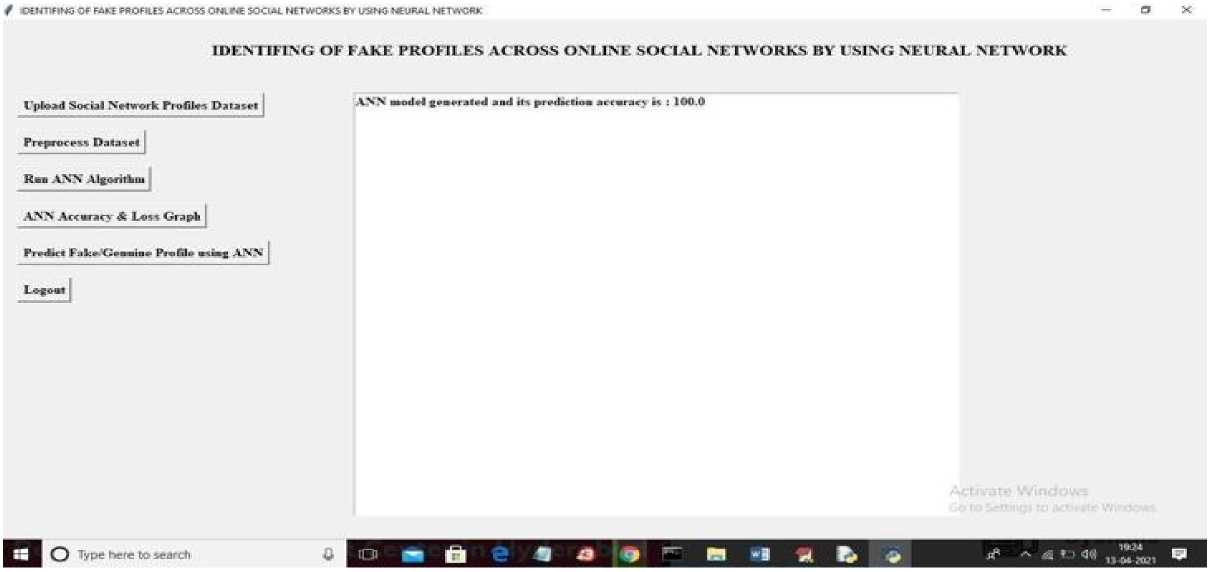
Binary

The output has only two values, either 0 and 1. For this, the threshold value is set up. If the net weighted input is greater than 1, the output is assumed as one otherwise zero.

Sigmoidal Hyperbolic

This function has an 'S' shaped curve. Here the tan hyperbolic function is used to approximate output from net input. The function is defined as - f (x) = (1/1+ exp(-????x)) where ???? - steepness parameter

IV. RESULTS



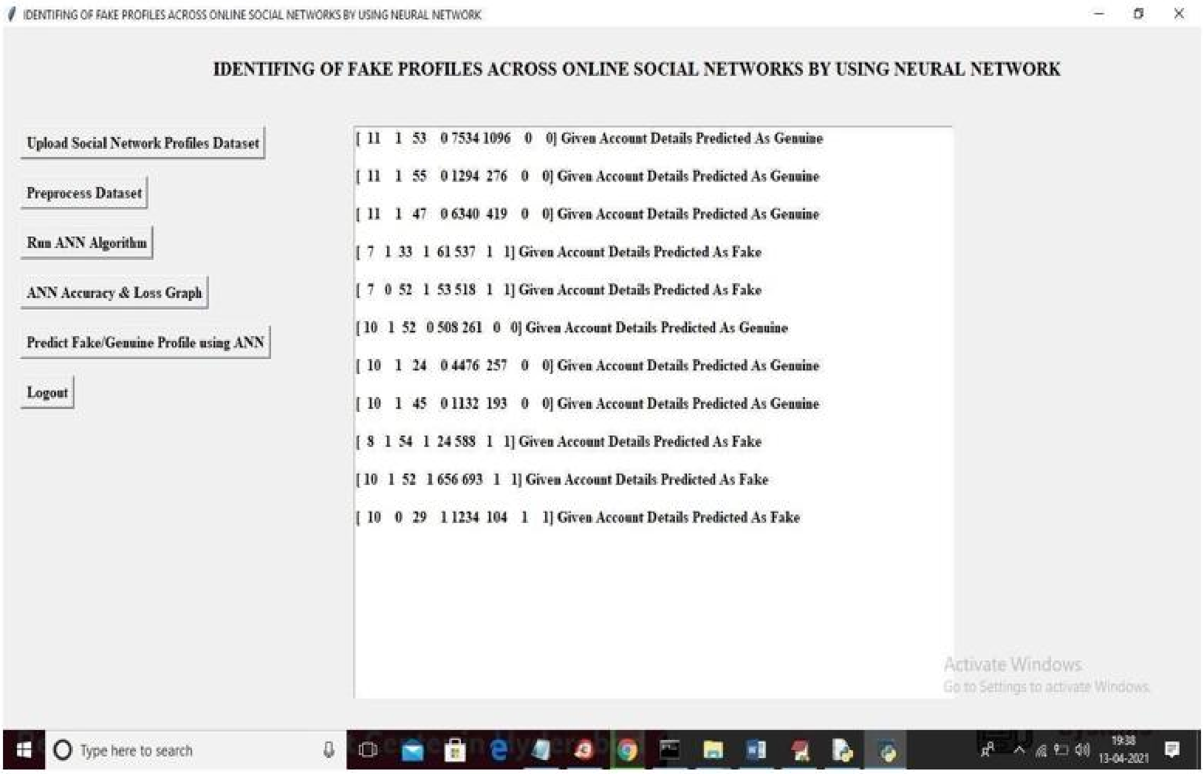
IDENTIFING OF FAKE PROFILES ACROSS ONLINE SOCIAL NETWORKS BY USING NEURAL NETWORK

IDENTIFING OF FAKE PROFILES ACROSS ONLINE SOCIAL NETWORKS BY USING NEURAL NETWORK

ANN model generated and its prediction accuracy is : 100.0

Upload Social Network Profiles Dataset
Preprocess Dataset
Run ANN Algorithm
ANN Accuracy & Loss Graph
Predict Fake/Genuine Profile using ANN
Logout
Activate Windows
Go to Settings to activate Windows.
Type here to search
19:24
13-04-2021

Fig. 4 Identification of fake facebook profiles page


IDENTIFING OF FAKE PROFILES ACROSS ONLINE SOCIAL NETWORKS BY USING NEURAL NETWORK
IDENTIFING OF FAKE PROFILES ACROSS ONLINE SOCIAL NETWORKS BY USING NEURAL NETWORK
Upload Social Network Profiles Dataset
Preprocess Dataset
Run ANN Algorithm
ANN Accuracy & Loss Graph
Predict Fake/Genuine Profile using ANN
Logout
Activate Windows
Go to Settings to activate Windows.
Type here to search
19:38
13-04-2021

```
[ 11   1  53   0 7534 1096   0   0] Given Account Details Predicted As Genuine
[ 11   1  55   0 1294  276   0   0] Given Account Details Predicted As Genuine
[ 11   1  47   0 6340  419   0   0] Given Account Details Predicted As Genuine
[ 7  1  33  1  61 537  1  1] Given Account Details Predicted As Fake
[ 7  0  52  1  53 518  1  1] Given Account Details Predicted As Fake
[10  1  52  0 508 261  0  0] Given Account Details Predicted As Genuine
[ 10   1  24   0 4476  257   0   0] Given Account Details Predicted As Genuine
[ 10   1  45   0 1132  193   0   0] Given Account Details Predicted As Genuine
[ 8  1  54  1  24 588  1  1] Given Account Details Predicted As Fake
[10  1  52  1 656 693  1  1] Given Account Details Predicted As Fake
[ 10   0  29   1 1234  104   1   1] Given Account Details Predicted As Fake
```

Fig. 5 Uploaded the social network profile dataset

V. CONCLUSION

In this project, we use deep learning, namely an artificial neural network to determine what are the chances that a friend request is authentic are or not. Each equation at each neuron (node) is put through a Sigmoid function. We use a training data set by Facebook or other social networks. This would allow the presented deep learning algorithm to learn the patterns of bot behavior by backpropagation, minimizing the final cost function and adjusting each neuron's weight and bias. In this project, we outline the classes and libraries involved. We also discuss the sigmoid function and how are the weights determined and used. We also consider the parameters of the social network page which are the most important to our solution.

REFERENCES

[1]https://www.statista.com/topics/1164/social-networks/

[2]https://www.cnbc.com/2018/01/31/facebook-earnings-q4-2017- arpu.html

[3]https://www.cnet.com/news/facebook-breach-affected-50-millionpeople/

[4]https://www.facebook.com/policy.php

[5]Qiang Cao, Michael Sirivianos, Xiaowei Yang, and Tiago Pregueiro. 2012.

Aiding the detection of fake accounts in large scale social online services. In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation (NSDI'12). USENIX Association, Berkeley, CA, USA, 15-15.

[6] Akshay J. Sarode and Arun Mishra. 2015. Audit and Analysis of Impostors: An experimental approachto detect fake profile in an online social network. In Proceedings of the Sixth International Conference on Computer and Communication Technology 2015 (ICCCT '15). ACM, New York, NY, USA, 1-8. DOI: https://doi.org/10.1145/2818567.2818568

[7] Devakunchari Ramalingam, Valliyammai Chinnaiah. Fake profile detection techniques in large-scale online social networks: A comprehensive review. Computers & Electrical Engineering, Volume65,2018,Pages165-177,ISSN0045-7906,https://doi.org/10.1016/j.compelecen g.2017.05.020.

[8]https://www.enigmasoftware.com/top-20-countries-the-most-cybercrime

[9]pages.cs.wisc.edu/~bolo/shipyard/neural/local.html

[10]https://stackoverflow.com/questions/40758562/can-anyone-explain-mestandardscaler

[11]https://pandas.pydata.org

[12]https://www.tutorialspoint.com/python_pandas/index.htm

[13]http://www.numpy.org

[14]https://www.mathworks.com/products/matlab.html

[15]http://www.deeplearning.net/software/theano/

[16] https://scikit-learn.org/stable/

[17] https://keras.io

[18] https://www.tensorflow.org